Running Head: CONSTRUCT VALIDITY OF THE POMS-A

Construct Validity of the Profile of Mood States - Adolescents for use with Adults

Peter C. Terry

University of Southern Queensland, Australia

Andrew M. Lane

University of Wolverhampton, United Kingdom

Gerard J. Fogarty

University of Southern Queensland, Australia

Revision submitted to *Psychology of Sport and Exercise*

3/10/01

Construct Validity of the Profile of Mood States - Adolescents for use with Adults

Abstract

Objectives: The purpose of the present study was to extend the validation of the Profile of Mood States-Adolescents (POMS-A: Terry, P. C., Lane, A. M., Lane, H. J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. Journal of Sports Sciences, 17, 861-872) from adolescent to adult populations.

Design: A strategy of assessing the invariance of the POMS-A factor structure among disparate samples and of testing relationships with concurrent measures was used.

Methods: The POMS-A was administered to 2,549 participants from four samples: Adult athletes prior to competition (n = 621), adult student athletes in a classroom (n = 656), adolescent athletes prior to competition (n = 676), and adolescent students in a classroom (n = 596). A subset of 382 adult student athletes was used to test the criterion validity of the POMS-A.

Results: Confirmatory factor analysis provided support for the factorial validity of a 24-item, six-factor model using both independent and multi-sample analyses. Relationships between POMS-A scores and previously validated measures, that were consistent with theoretical predictions, supported criterion validity.

Conclusion: Supporting evidence was found that the psychometric integrity of the POMS-A extended from adolescent to adult populations.

Keywords: Measurement; Sport; Structural equations; Model testing; Emotion; Affect

Construct validity of the POMS-A for use with adults

The psychology of mood and sport performance has been researched extensively over the past thirty years (see LeUnes & Burger, 1998; LeUnes, 2000). Research investigating mood in sport has typically used the Profile of Mood States (POMS: McNair, Lorr, & Droppleman, 1971), a 65-item inventory that assesses six dimensions of the mood construct: anger, confusion, depression, fatigue, tension, and vigour. The brevity required of measures of psychological states in some research environments has spawned many shortened versions of the POMS (e.g., Grove & Prapavessis, 1992; McNair et al., 1992; Shacham, 1983; Terry, Lane, Lane, & Keohane, 1999).

Of the shortened versions, the Profile of Mood States-Adolescents (POMS-A: Terry et al., 1999) has been subjected to perhaps the most rigorous validation process. The POMS-A was developed in three stages. Stage 1 established content validity, whereby a panel of experts assessed an initial item pool for comprehensibility by adolescents and a sample of adolescents identified those items that best described each mood dimension. In Stage 2, a 24-item, six-factor structure was tested using confirmatory factor analysis on adolescents in a classroom setting and adolescent athletes before competition. The hypothesised model was supported in both groups independently and simultaneously. In Stage 3, relationships between POMS-A scores and previously validated measures, that were consistent with theoretical predictions, supported criterion validity.

Given that the POMS-A has shown encouraging psychometric properties among samples of adolescents, the question arises of whether the measure is equally suitable for use with adults. Conceptually, the answer appears straightforward. The adaptation of the adolescent version from the original adult version (McNair et al., 1971) involved the selection of mood descriptors that were judged to be age-appropriate for adolescents. In essence, the adolescent measure is a simpler and shorter version of the adult measure.

There appears to be no compelling reason why the adolescent version would not remain appropriate for use with adults. Indeed, for some adult populations, such as those with intellectual disabilities or limited education, a measure using simple language would seem highly appropriate, with potentially wide-ranging applications. Also, there is nothing in the small body of research literature that has investigated the link between age and mood (e.g., McNeil, Stone, Kozma and Andres, 1994) to suggest that adolescents conceptualise the notion of mood any differently to adults. Nevertheless, Comrey (1988) emphasized that the establishment of factorial validity is a necessary pre-requisite to the use of any scale in a second population and therefore, from a psychometric perspective, it is important to address the question of whether the measurement model for the POMS-A that was supported among adolescent samples would also be supported among adults. This question provided the central focus for the present study.

Schutz (1994) has argued that the proliferation of psychometric tests that have not been rigorously validated has led to conceptual confusion about some constructs investigated by sport psychologists. Watson and Tellegen (1985) raised a similar point in their review of mood measures. If the validity of a measure, or the appropriateness of its use in a particular environment, is in doubt then it is not possible to accurately test a related theory. Given that the POMS-A[1] has been used recently with adult samples for both research (e.g., Lane, Terry, Beedie, Curry, & Clark, 2001; Terry, Carron, Pink, Lane, Jones, & Hall, 2000; Terry, Lane, & Warren, 1999) and applied purposes (e.g., Lane & Terry, 1998; Vleck, Garbutt, & Terry, 1998), it appears that evidence of the validity of the scale for use with adults is overdue.

Although tests of factorial validity provide evidence that items in a factor assess the

[1] The measure reported in some of these studies was the POMS-C (Terry, Keohane, & Lane, 1996). The name of the questionnaire was changed to POMS-A following the review process of the validation study.

same underlying construct, the researcher decides what the factor is called. Therefore, an important step in establishing the validity of a questionnaire is to show that it actually measures the construct it purports to measure. Tests of criterion validity help to clarify the meaning of measured constructs by assessing relationships with other measures against theoretical predictions. Although Terry et al. (1999) tested the POMS-A against three related scales, it is important to test the scale further against criterion measures, especially since the measure is now being applied to a different population. Therefore, the second purpose of the present study was to assess the criterion validity of the POMS-A among adult participants.

Considering the proposed importance of a strong theoretical underpinning for psychological tests (Schutz, 1994) it is incumbent upon researchers to establish the theoretical integrity of measures. Extensive conceptual discussion of the mood construct generally, and the model underlying the POMS-A in particular, have been presented elsewhere (Lane & Terry, 2000; McNair et al., 1971; Terry, in press). It is beyond the scope of the present paper to reproduce those theoretical discussions, but to facilitate an appreciation of the proposed item groupings, a general description of the six mood dimensions is presented.

Terry et al. (1999) described the factors of the POMS-A in the following way, "Anger is typified by feelings that vary in intensity from mild annoyance or aggravation to fury and rage, and is associated with arousal of the autonomic nervous system (Spielberger, 1991). Confusion is proposed to be a feeling state characterised by bewilderment and uncertainty, associated with a general failure to control attention and emotions. Depression is associated with a negative self-schema characterised by themes such as hopelessness, personal deficiency, worthlessness, and self-blame (Beck & Clark, 1988). Fatigue is typified by feelings of mental and physical tiredness. Tension is typified by feelings such as nervousness, apprehension, worry, and anxiety. Vigour is typified by feelings of excitement, alertness, and physical energy" (p.863). Based on the findings of Terry et al., it was hypothesised that, in the present study, depression would show moderate positive relationships with anger, confusion, fatigue and tension, and a weak inverse relationship with vigour; while vigour would show a moderate inverse relationship with fatigue but be unrelated to anger, confusion, and tension.

Method

1. Model Testing

Participants

A multi-sample approach was used to test the invariance of the factor structure of the POMS-A among four disparate samples. Given that the central purpose of the study was to test whether the factor structure remained the same among adults and adolescents, two adult samples and two adolescent samples were recruited. Sample 1 comprised 621 adult athletes (463 males, 158 females; age: M = 27.2 yr., SD = 6.6 yr.), recruited from the sports of cycling, distance running, kickboxing, rowing, and swimming, to reflect a wide range of age, experience, and ability. Sixty-five cyclists (age: M = 29.6 yr., SD = 7.6 yr.) with a mean of six years of racing experience were drawn from a 10-mile time trial; 297 distance runners (age: M = 31.0 yr., SD = 10.1 yr.) with a mean of six years of racing experience were drawn from cross-country, 10 k, 6 miles and marathon events; 89 kickboxers (age: M = 24.6 yr., SD = 4.8 yr.) were drawn from the non-contact, semi-contact and full contact categories at a national championship; 98 rowers (age: M = 23.3 yr., SD = 7.2 yr.) were drawn from the 1996 world championships; and 72 swimmers (age: M = 20.5 yr., SD = 2.1 yr.) with a mean of seven years of racing experience were drawn from club events. Sample 2 comprised 656 adult students studying for a degree in sport sciences or a qualification in fitness training and leisure (365 males, 291 females; age: M = 24.2 yr., SD

= 3.7 yr.). Participants in Sample 2 were from the sports of basketball, boxing, duathlon, distance running, hockey, karate, rugby league, rugby union, soccer, swimming, taekwondo, tennis, track and field and triathlon. Sample 3 comprised 676 adolescent athletes (301 males, 375 females; age: M = 14.7 yr., SD = 1.8 yr.) competing in the London Youth Games, from the sports of archery, hockey, judo, netball, soccer, table tennis, track and field, trampolining, triathlon and volleyball. Sample 4 comprised 596 adolescent students (313 males, 283 females; age: M = 14.7 yr., SD = 1.4 yr.) from secondary schools in the west London area.

Procedures

Given the wide range of potential uses to which a simple measure of mood can be applied, it is appropriate that tests of validity are conducted in more than one setting. Investigations of mood responses among athletes have often been conducted at the pre-competition stage, but also in situations away from the competition environment (see LeUnes, Hayward, & Daiss, 1988). Normative data provided by Terry and Lane (2000) demonstrated that, typically, mood responses vary between competitive and non-competitive situations. Therefore, it was decided to assess moods in both situations. Using this strategy, it was possible to determine whether the factor structure of the POMS-A remained invariant across adolescent and adult samples even in situations that differed in degree of ego involvement. Participants in Sample 1 (adult athletes) and Sample 3 (adolescent athletes) completed the POMS-A approximately one hour before a competition. Participants in Sample 2 (adult students) and Sample 4 (adolescent students) completed the POMS-A at the start or end of a class. All participants were asked to rate, “How are you feeling right now?” in terms of the 24 mood descriptors, e.g., “alert”, “unhappy”. The POMS-A has a five-point response scale, from 0 (not at all) to 4 (extremely). The instructions to participants included a reminder to respond to all items and a statement designed to discourage a social desirability bias (c.f., Martens, Vealey, & Burton, 1990). A culturally appropriate, alternative word list (c.f., Albrecht & Ewing, 1989) was made available to participants for reference in case mood descriptors could not be understood.

Following the recommendations of Byrne (2000), the hypothesized 24-item, six-factor model of mood was first tested on each sample independently before conducting a multi-sample analysis, in which the hypothesized model was tested on all four samples simultaneously. Confirmatory factor analysis (CFA) using EQS V5 (Bentler & Wu, 1995) was used to test the model, which specified that items were related to their hypothesised factor with the variance of the factor fixed at 1. Consistent with theoretical predictions and previous empirical support, the latent factors anger, confusion, depression, fatigue, and tension were allowed to correlate (see Terry et al., 1999). Vigour was allowed to correlate with depression and fatigue only, as it had been hypothesised that relationships between vigour and anger, confusion, and tension would not differ significantly from zero.

The choice of cut-off criteria used to evaluate model adequacy is a contentious issue. Some researchers favour a two-index strategy, with the indices selected on the basis of sample size, model complexity, and the distributional properties of the data (Hu & Bentler, 1999). We followed the approach of Byrne (1998, 2000), Hoyle and Panter (1995) and Kline (1998), all of whom advocated use of a range of fit indices to judge model adequacy. Kline, for example, recommended a “minimal set” that included (a) the χ^2 statistic and its associated degrees of freedom, (b) an index that describes the overall proportion of variance explained, (c) an index that adjusts the proportion of variance explained for model complexity, and (d) an index based on the standardized residuals.

We chose a four-index strategy. The first index used to judge model adequacy was the ratio of χ^2 to degrees of freedom. There is disagreement about what size of ratio indicates an acceptable fit, with estimates varying from two to five. Kline (1998) proposed that a ratio of less than three is acceptable. Two incremental fit indices were used; the comparative fit index (CFI: Bentler, 1990) and the non-normed fit index or Tucker-Lewis index (TLI: Tucker & Lewis, 1973). Incremental fit indices are based on comparisons between the hypothesised model and a null model (in which there are no relationships among the observed variables) and are not influenced by sample size. Kline (1998) proposed that values for the CFI and TLI of less than .90 indicate that the hypothesized model could be substantially improved, whereas Hu and Bentler (1999) suggested that, in most circumstances, values should approach .95. The fourth index used was the root mean square error of approximation (RMSEA: Steiger, 1990), which indicates the mean discrepancy between the observed covariances and those implied by the model per degree of freedom, and therefore has the advantage of being sensitive to model complexity. A value of .05 or lower indicates a good fit and values up to .08 indicate an acceptable fit (Browne & Cudeck, 1993). Byrne (1998) described the RMSEA as “one of the most informative criteria in structural equation modelling” (p. 112).

Multi-sample CFA was used to test the strength of the factor solution in all four samples simultaneously. In multi-sample analysis, it is assumed that data from more than one sample provide comparable information about the hypothesised model. This assumption is tested by analysing data from different samples simultaneously to verify whether the model reproduces the data of each sample to within sampling accuracy (Bentler, 1995). Bentler recommended that hypothesis testing in multi-sample analysis should be sequential. Hence, the first hypothesis was that the factor solution would remain the same in all four samples. This hypothesis was tested with no equality constraints in place, to establish a baseline against which to compare subsequent, more constrained, models. The second hypothesis tested was that factor loadings would remain invariant in all four samples. The third hypothesis was that factor loadings and inter-correlations among factors would remain invariant in all four samples.

2. Test of Criterion Validity

Participants

A total of 382 adult student athletes completed the POMS-A and a second questionnaire. Ninety-one participants (Age: M = 21.0 yr. SD = 5.2 yr.) completed the original POMS (McNair et al., 1971), 84 participants (Age: M = 23.3 yr., SD = 3.5 yr.) completed the Positive and Negative Affect Schedule (PANAS: Watson, Clark, & Tellegen, 1988), 97 participants (Age: M = 23.5 yr.; SD = 3.4 yr.) completed the state anger scale of the State-Trait Anger-expression Inventory (STAXI: Spielberger, 1991); and 110 participants (Age: M = 25.9 yr., SD = 10.4 yr.) completed the depression scale of the Hospital Anxiety and Depression Scale (HADS: Zigmond & Snaith, 1983). Participants completed the questionnaires in accordance with the procedure used for the student athletes in Stage 1.

Criterion Measures

Two considerations are particularly salient when choosing appropriate criterion measures. First, a criterion scale should itself be a valid, reliable measure. Second, it should be possible to predict the relationship between scores on the measure being validated and the criterion measure. All the criterion measures selected meet these two considerations. The original POMS was an obvious choice as it assesses the same six mood dimensions as the POMS-A. The PANAS was selected because it assesses two broad affective dimensions that are conceptually related in a predictable way to the POMS-A scales. The STAXI and the HADS were selected because they assess specific constructs that form part of the POMS-A and therefore should show strong relationships with some scales of the POMS-A but not with others. It is acknowledged that these arguments would apply equally to other potential criterion measures.

Profile of Mood States (McNair et al., 1971)

The 65-item POMS was developed via six factor analytic studies. McNair et al. showed evidence of concurrent and predictive validity, and produced normative data for students and psychiatric outpatients. McNair et al. (1992) claimed that the POMS was valid for use in sport and exercise environments and provided a summary of findings from these domains in support of this proposition. A response set of "How are you feeling right now" was used in the present study. Given that the POMS-A is a derivative of the original POMS, strong positive relationships were hypothesised between the respective anger, confusion, depression, fatigue, and tension scales. A moderate positive relationship was hypothesised between the two vigour scales because the original scale assesses a fairly broad-based positive mood (including items such as "cheerful" and "carefree") whereas the POMS-A vigour scale assesses a narrower construct ("active", "alert", "energetic", "lively").

Positive and Negative Affect Schedule (Watson et al., 1988)

Watson et al. developed the 20-item PANAS to assess independent markers of positive and negative affect. Validation studies for the PANAS involved 3,554 applications of the scale. The two factors showed strong content validity, with all items loading at .50 or higher onto their hypothesised factor. PANAS items are rated on a 5-point scale, from 1 (not at all) to 5 (extremely). Examples of positive affect items include "excited", "enthusiastic" and "determined." Examples of negative affect items include "distressed", "guilty" and "scared." Recent research has supported the factor structure of the PANAS among a sample of young athletes (Crocker, 1997). In the present study, it was hypothesised that the POMS-A vigour scale would correlate with the positive affect scale of the PANAS, while the POMS-A tension, depression, anger, fatigue, and confusion scales would correlate with the PANAS negative affect scale.

<u>State-Trait Anger-expression Inventory</u> (Spielberger, 1991)

The 10-item state anger scale was validated on a sample of 550 participants. Exploratory factor analysis identified a single factor with an alpha coefficient (Cronbach, 1951) of .93. STAXI items are rated on a 4-point scale, from1 (almost never) to 4 (very often). In the present study, it was hypothesised that the POMS-A anger scale would be highly correlated with STAXI scores, whereas the other five POMS-A scales would be unrelated or moderately correlated with the STAXI scale.

<u>Hospital Anxiety and Depression Scale</u> (Zigmond & Snaith, 1983)

The HADS includes seven items to assess depression among medical outpatients and the general population. Items are rated on a 4-point scale, from 0 (not at all, or only occasionally) to 3 (most of the time, or a great deal of the time). Validation of the HADS used 100 outpatients who completed the scale as part of a psychiatric interview. The authors reported an alpha coefficient of .60 for the depression scale. Concurrent validity was assessed against information given in a 20-minute interview to an interviewer blind to the HADS scores. This technique produced a correlation coefficient of .79 for the depression scale. The HADS anxiety scale was not used in the present study. Given the proposed pivotal position of depressed mood in mood-performance relationships (see Lane & Terry, 2000), the criterion validity of the POMS-A depression scale was of particular interest. As depressed mood was assessed using two different response sets (“How are you feeling right now?” for the POMS-A and “How have you been feeling during the past week?” for the HADS) it was hypothesised that the two scales would be moderately, rather than highly correlated. It was also hypothesised that the other five POMS-A scales would be unrelated or be weakly correlated with the HADS depression scale.

Results

Single-sample Confirmatory Factor Analysis

Prior to analysis, each data set was screened to check that the assumptions of univariate and multivariate normality had been met. Evidence of non-normality was found for some variables. Inspection of cases identified as outliers suggested that although the response patterns for these individuals were unusual, they were nevertheless plausible, so no attempt was made to transform variables or to trim the data set. To compensate for the non-normality, the Satorra-Bentler χ^2, a statistic that includes a downward correction for degree of observed kurtosis (Satorra & Bentler, 1994), was used to test the model fit for individual groups. This strategy was supported in a recent review of research methods in sport and exercise psychology (Biddle, Markland, Gilbourne, Chatzisarantis, & Sparkes, 2001).

Overall, fit indices showed moderate support for the tenability of the hypothesised model (see Table 1). Generally, the four groups showed similar fit, although the indices were somewhat lower for the adolescent student group. Modification indices suggested a number of ways by which fit could be improved, but only one change was implemented. This change allowed the error terms for the fatigue indicators “sleepy” and “tired” to covary. A χ^2-difference test showed that this modification significantly improved fit for all groups but had the greatest impact on the fit indices for the adolescent athletes (CFI = .936, TLI = .926, RMSEA = .054) and adolescent students (CFI = .936, TLI = .948, RMSEA = .046). From a theoretical point of view, it made sense to allow these terms to covary because the presence of unique variance within the fatigue factor suggested that the adolescent samples had difficulty separating the meaning of “sleepy” and “tired”.

As shown in Table 1, even without this modification, the χ^2 /df ratios met the criterion value for a good fit in all samples, RMSEA values were acceptable, while the CFI and TLI values fell between traditional and recent benchmarks. Factor loadings were all highly significant. More importantly, the magnitude of the factor loadings (72% were above .70) further supported the validity of the factor structure (see Table 2).

Correlations among mood dimensions are contained in Table 3. The direction and magnitude of relationships was consistent with the hypothesised model and with those reported by Terry et al. (1999). The Lagrange multiplier test indicated that the fit of the model would not be improved by allowing vigour to correlate with anger, tension and confusion. Collectively, single-sample CFA results provided support for the tenability of the hypothesised model, and hence the next step was to examine the congruency of the hypothesised model against data in the four samples simultaneously, using multi-sample CFA.

Multi-sample Confirmatory Factor Analysis

The results of the multi-sample CFA are contained in Table 4. The fit of the baseline model with no equality constraints imposed was encouraging whereas the fit of the more constrained models was marginal. Modification indices suggested a number of potential changes to all models. At this point, the researchers were confronted with the task of identifying those parts of the model that needed re-specification to achieve better fit. However, problems associated with post-hoc modifications have been emphasized recently (Biddle et al., 2001) and, indeed, specialists in structural equation modelling (e.g., Gerbing & Hamilton, 1996) have pointed out that post-hoc modifications are more exploratory than confirmatory. In fact, exploratory factor analysis (EFA) is considered a viable alternative to attempting to adjust the confirmatory model. Gorsuch (1997) proposed that in situations where model fit is difficult because of many small deviations from the hypothesized model, a situation we were faced with in the present analysis, EFA is “an appropriate alternative to attempting to adjust the confirmatory model” (p. 536).

Mindful of this advice, and mindful also of advice from other researchers (e.g., Comrey, 1988; Gerbing & Hamilton, 1996) that CFA and EFA should be used together in scale construction and validation, we decided to use EFA with each of the groups separately to see how the items loaded when totally unconstrained, apart from the stipulation that the number of factors must equal six. The maximum likelihood method was used with oblique rotation.

When EFA is used for confirmatory purposes, the expectation is that the hypothesised factor structure of the scale will be recovered in all samples, just as it would be in a CFA (Gorsuch, 1997). In the case of the POMS-A, the expectation was that each of the 24 items should have a large loading on its associated factor and negligible loadings (cross-loadings) on all other factors. By convention, factor loadings below .30 are usually considered to be non-significant (e.g., Tabachnick & Fidell, 1996). In the present analyses, we adopted a more rigorous criterion of .20. In other words, cross-loadings of .20 or more were counted as evidence of misfit and likely sources of the model misspecification in the CFA section of our analyses.

The results showed that for three of the four samples, the expected factor pattern emerged. That is, all six factors emerged clearly with each factor defined solely by appropriate marker items. Explained variance for the four samples ranged from 56% to 63%. Furthermore, the patterns for three of the samples supported the hypothesized model. For the adult athletes, only the item “uncertain” cross-loaded above .20, with a loading of .31 on tension and .51 on confusion. For the adult students, an identical pattern emerged with “uncertain” loading on both tension (.36) and confusion (.39) but no other cross-loadings above .20. For the adolescent athlete group, three cross-loadings were found. The item “downhearted” loaded on anger (.22) and depression (.56); “muddled” loaded on depression (.33) and confusion (.54); and “uncertain” loaded on tension (.28) and confusion (.54). Thus, for these three groups, apart from a small number of cross-loading items, the derived factor structure was exactly as hypothesized.

Some variation to this pattern was observed with the adolescent student sample. As was the case with the two adult samples, “uncertain” was the only item with a cross loading (.25 on tension and .53 on confusion). However, the fatigue dimension split into two correlated ($\underline{r}$ = .51) factors with “exhausted” and “worn out” defining one factor and “sleepy” and “tired” the other. This tendency had already been detected in the CFA and was modelled by fitting a covariance pathway to the error terms for “sleepy” and “tired”. To compensate for this additional factor, the analysis was re-run, this time requesting seven factors. When this was done, with the exception of the cross loading for the item “uncertain”, the hypothesized factors emerged for anger, confusion, depression, tension and vigour, plus two fatigue factors of two items each.

Overall, the combined results of the confirmatory and exploratory analyses add weight to the notion that the hypothesised factor structure of the POMS-A can be reproduced in disparate samples, thereby supporting the factorial validity of the measure.

Criterion Validity

Relationships between scores on the POMS-A and criterion measures are contained in Table 5. Correlations between the POMS-A and the original POMS support the notion that, generally, the scales measure the same thing. The moderate relationship between the two vigour scales suggests that they assess slightly different constructs. The POMS-A vigour scale, which includes the items “active”, “alert”, “energetic” and “lively”, may provide a more focused measure of the vigour construct than the original POMS scale, which includes items such as “cheerful” and “carefree”.

The strong relationship between the depression scales of the POMS-A and the original POMS suggests that they assess a very similar construct. Kline (1998) cautioned that reducing the number of items in a scale might yield a collection of items almost identical in meaning, thereby boosting internal consistency. The present results are consistent with the notion that the POMS-A depression scale of four items assesses essentially the same construct as the original POMS depression scale, which includes 15 items. The POMS-A depression scale and the HADS depression scale were, as hypothesized, moderately correlated.

It should be noted that the POMS-A provides a measure of depressed mood at a given point in time not a measure of clinical depression. For clinical depression, Tennen, Hall, and Affleck (1995) proposed that self-report measures should be used in conjunction with follow-up interviews. Therefore, the validity of the POMS-A for use with clinical populations is unknown.

As hypothesised, PANAS scores for positive affect showed a strong positive correlation with scores on the POMS-A vigour scale but minimal correlation with the other POMS-A scales. Also as hypothesised, PANAS scores for negative affect correlated with POMS-A scores for anger, confusion, depression, fatigue and tension but were unrelated to scores for vigour. Further, STAXI scores correlated strongly with POMS-A scores for anger. Overall, correlations between scores on the POMS-A, the PANAS, and the STAXI were consistent with those reported by Terry et al. (1999). Collectively, it is proposed that the pattern of correlations between scores on the POMS-A and criterion measures provides strong evidence of concurrent validity.

Discussion

The purpose of the present study was to validate the POMS-A for use with adult athletes. Individual CFAs suggested that the measurement model underlying the POMS-A provided an adequate fit to the data for each of the four samples in this study. Lagrange multiplier (LM) tests indicated a number of ways in which fit could be improved.

However, bearing in mind that the large samples used in this study rendered the LM tests particularly powerful, we would argue that although statistically significant, many of these differences are not of great practical or theoretical importance.

The one concession to the LM tests was the fitting of a pathway between the error terms for “tired” and “sleepy”. This modification was suggested for all samples but it had its greatest impact on the fit statistics for the adolescent students. The inference we draw from this modification is that detection of the semantic gap between “tired” and “sleepy” is more problematic for adolescents than adults. In other words, adolescents may have difficulty distinguishing between these two mood descriptors. Context is important here. In a sport setting, "tired" and "sleepy" have very different meanings whereas in a classroom setting, these two descriptors could mean almost the same thing. This reinforces our strategy of checking for consistency across settings as well as across age groups. Having items with such an overlap does not damage the psychometric properties of the instrument but, as we have demonstrated, the overlap needs to be modelled. Solutions to this problem would include dropping one of the items, combining them to form a parcel or, as we recommend, allowing their error terms to covary.

A multi-sample CFA provided some support for the notion that the factor structure of the POMS-A, the factor loadings of the items, and the inter-correlations among subscales did not vary greatly in the four different samples. Fit statistics were judged to be adequate when factor structure was constrained to be equal across samples, but were considered marginal when factor loadings and factor covariances were also constrained to be equal. Again, LM tests indicated a number of potential modifications. To assist in identifying potentially important differences among samples, EFA was used. The results of the EFAs helped to clarify the differences between samples. They demonstrated that the structure of the POMS-A was easily recovered for all samples, and showed that the cross-loadings that caused some of the misfit in the CFAs were relatively minor. Indeed, they would have been ignored had EFA been used initially to assess the factor structure of the POMS-A. The EFAs also highlighted a problem that has already been noted, that is, the difficulty of trying to fit a single factor to the markers for fatigue. It is problematic for the adolescent students because of the conceptual overlap between the items “tired” and “sleepy”. Finally, the EFAs suggested that for all samples, but especially the older ones, allowing the item “uncertain” to load on both tension (minor loading) and confusion (major loading) would improve the measurement model.

In summary, these findings indicated that factorial validity was generally supported among disparate samples and the hypothesised relationships among mood dimensions were demonstrated. It should be noted that model fit was actually stronger among adults than among adolescents. Collectively, findings supported the notion that the POMS-A has shown acceptable indicators of validity as a measure of mood among adults as well as adolescents.

Criterion validity was supported by the relationships between POMS-A scores and other measures taken concurrently. Comrey (1988) stressed the importance of developing a nomothetic network of evidence supporting claims that the instrument measures what it is intended to measure. In the words of Anastasi: “It is only through the empirical investigation of the relationship of test scores to other external data that we can discover what a test measures” (cited in Comrey, 1988). In a sense, the validation of a test never ends. In the present study, we have demonstrated that the constructs measured by POMS-A have predictable relationships with positive and negative affect as measured by PANAS, anger as measured by the STAXI, depression as measured by the HADS, and mood states as measured by the original version of the POMS.

It is suggested that the POMS-A is an appropriate tool with which to test mood theory. Importantly, the brevity of the POMS-A facilitates mood assessment in research environments where there is a limited amount of time available for data collection. One such line of investigation is the assessment of mood before sport competition. Examination of the mood and performance relationship has been prominent in sport psychology (see LeUnes & Burger, 1998; LeUnes, 2000 for reviews) and recent research has started to examine the mechanisms underlying mood-performance relationships (Lane & Terry, 2000).

In conclusion, the purpose of the study was to extend the validation of a measure of mood from adolescent to adult samples. Confirmatory procedures offered support for the tenability of a six-factor model using both independent and multi-sample analyses. Criterion validity was supported via relationships with previously validated inventories. Overall, it is suggested that the construct validity of the POMS-A has been shown to be satisfactory, and therefore the scale may provide a useful measure of mood for future research or applied work.

References

Albrecht, R.R., & Ewing, S.J. (1989). Standardizing the administration of the Profile of Mood States (POMS): Development of alternative word lists. Journal of Personality Assessment, 53, 31-39.

Beck, A.T., & Clark, D.A. (1988). Anxiety and depression: An information processing perspective. Anxiety Research, 1, 23-56.

Bentler, P.M. (1990). Comparative fit indexes in structural models. Psychological Bulletin, 107, 238-246.

Bentler, P.M. (1995). EQS Structural equation program manual. Los Angeles, CA: BMDP Statistical Software.

Bentler, P.M., & Wu, E.J.C. (1995). EQS/Windows user's guide. Los Angeles, CA: BMDP Statistical Software.

Biddle, S.J.H., Markland, D., Gilbourne, D., Chatzisarantis, N.L.D., & Sparkes, A.C. (2001). Research methods in sport and exercise psychology: quantitative and qualitative issues. Journal of Sports Sciences, 19, 777-809.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K.A., & Long, J.S. (Eds.), Testing structural equation models, (pp.132-162). Newbury, CA: Sage.

Byrne, B.M. (1998). Structural equation modelling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications and programming. Mahwah, NJ: Lawrence Erlbaum.

Byrne, B.M. (2000). Structural equation modelling with AMOS: Basic concepts, applications and programming. Mahwah, NJ: Lawrence Erlbaum.

Comrey, A.L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. Journal of Consulting and Clinical Psychology, 56, 754-761.

Crocker P.R.E. (1997). A confirmatory factor analysis of the Positive and Negative Affect Schedule (PANAS) with a youth sport sample. Journal of Sport and Exercise Psychology, 19, 91-97.

Cronbach, L.J. (1951). Coefficient alpha and internal structure of tests. Psychometrika, 16, 297-334.

Gerbing, D.W., & Hamilton, J.G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. Structural Equation Modeling, 3, 62-72.

Gorsuch, R.L. (1997). Exploratory factor analysis: Its role in item analysis. Journal of Personality Assessment, 68, 532-560.

Grove, J.R., & Prapavessis, H. (1992). Preliminary evidence for the reliability and validity of an abbreviated Profile of Mood States. International Journal of Sport Psychology, 23, 93-109.

Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In Hoyle, R. H. (Ed.), Structural equation modeling: Concepts, issues, and applications, (pp. 158-175). Newbury, CA: Sage.

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modelling, 6, 1-55.

Kline, P. (1998). Principles and practice of structural equation modelling. New York: The Guildford Press.

Lane, A.M., & Terry, P.C. (1998). Development of normative data for the Profile of Mood States-C among adult and young athletes [Abstract]. Journal of Sports Sciences, 16, 95-96.

Lane, A.M., & Terry, P.C. (2000). The nature of mood: Development of a conceptual model with a focus on depression. Journal of Applied Sport Psychology, 12, 16-33.

Lane, A.M., Terry, P.C., Beedie, C.J., Curry, D.A., & Clark, N. (2001). Mood and performance: Test of a conceptual model with a focus on depressed mood. Psychology of Sport and Exercise, 2,

LeUnes, A. (2000). Updated bibliography on the Profile of Mood States in sport and exercise psychology research. Journal of Applied Sport Psychology, 12, 110-113.

LeUnes, A., & Burger, J. (1998). Bibliography on the Profile of Mood States in sport and exercise, 1971-1995. Journal of Sport Behavior, 21, 53-70.

LeUnes, A., Hayward, S.A., & Daiss, S. (1988). Annotated bibliography on the Profile of Mood States in sport, 1975-1988. Journal of Sport Behavior, 11, 213-240.

Martens, R., Vealey, R.S., & Burton, D. (1990). Competitive anxiety in sport. Champaign, IL: Human Kinetics.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1971). Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Services.

McNair, D.M., Lorr, M., & Droppleman, L.F. (1992). Revised manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Services.

McNeil, J.K., Stone, M.J., Kozma, A., & Andres, D. (1994). Age differences in mood: Structure, mean level, and diurnal variation. Canadian Journal on Aging, 13, 201-220.

Satorra, A. & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), Latent variables analysis (pp. 399-419). Thousand Oaks, CA: Sage.

Schutz, R.W. (1994). Methodological issues and measurement problems in sport psychology. In S. Serpa, J. Alves, & V. Pataco (Eds.), International perspectives on sport and exercise psychology, (pp. 35-57). Morgantown, WV: Fitness Information Technology.

Shacham, S. (1983). A shortened version of the Profile of Mood States. Journal of Personality Assessment, 47, 305-306.

Spielberger, C.D. (1991). Manual for the State-Trait Anger-expression Inventory. Odessa, FL: Psychological Assessment Resources.

Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. Multivariate Behavioral Research, 25, 173-180.

Tabachnick, B.G., & Fidell, L.S. (1996). Using multivariate statistics (3rd ed.). New York, NY: Harper and Row.

Tennen, H., Hall, J., & Affleck, G. (1995). Depression research methodologies in the Journal of Personality and Social Psychology: A review and critique. Journal of Personality and Social Psychology, 68, 870-884.

Terry, P.C. (in press). Mood and emotions in sport. In T. Morris, & J. Summers (Eds.), Sport psychology: Theory, applications and issues. Brisbane: Wiley.

Terry, P.C., Carron, A.V., Pink, M.J., Lane, A.M., Jones, G., & Hall, M. (2000). Team cohesion and mood in sport. Group Dynamics: Theory, Research and Practice, 4, 244-253.

Terry, P.C., Keohane, L., & Lane, H.J. (1996). Development and validation of a shortened version of the Profile of Mood States suitable for use with young athletes [Abstract]. Journal of Sports Sciences, 14, 49.

Terry, P.C. & Lane, A.M. (2000). Development of normative data for the Profile of Mood States for use with athletic samples. Journal of Applied Sport Psychology, 12, 69-85.

Terry, P.C., Lane, A.M., Lane, H.J., & Keohane, L. (1999). Development and validation of a mood measure for adolescents. Journal of Sports Sciences, 17, 861-872.

Terry, P.C., Lane, A.M., & Warren, L. (1999). Eating attitudes, body shape perceptions, and mood among elite rowers: Effects of age, gender and weight category. Journal of Science and Medicine in Sport, 2, 67-77.

Tucker, L.R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Watson, D., & Tellegen, A. (1985). Toward a conceptual structure of mood. Psychological Bulletin, 98, 219-235.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Vleck, V., Garbutt, G., & Terry, P.C. (1998). Development of triathlon-specific normative data for the Profile of Mood States-C [Abstract]. Journal of Sports Sciences, 16, 399.

Zigmond, A.S., & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. Acta Psychiatrica Scandinavia, 67, 361-370.

Author Note

Peter C. Terry and Gerard J. Fogarty, Department of Psychology, University of Southern Queensland, Australia.

Andrew M. Lane, School of Sport, Performing Arts, and Leisure, University of Wolverhampton, United Kingdom.

The contribution of Anthony Darroch, Valerie Darroch, Susan Firth, Lee Keohane, Helen Lane, John Lawson, and Lucinda Warren during the data collection and/or processing stages is gratefully acknowledged.

Correspondence concerning the article should be addressed to Peter C. Terry, Department of Psychology, University of Southern Queensland, Toowoomba, Queensland, 4350, Australia, or by email to terryp@usq.edu.au

Table 1

Confirmatory Factor Analysis of the POMS-A with Adult and Adolescent Samples

Fit Indices	Sample			
	Sample 1 (n = 621)	Sample 2 (n = 656)	Sample 3 (n = 676)	Sample 4 (n = 596)
χ^2 (df = 240)	524.761	521.484	424.785	373.800
χ^2:df ratio	2.203	2.191	1.783	1.574
CFI	.923	.941	.914	.930
TLI	.912	.932	.901	.919
RMSEA	.059	.061	.062	.057

Note. All χ^2 values are significant at $p < .01$.

Sample 1 = adult athletes, Sample 2 = adult students, Sample 3 = adolescent athletes, Sample 4 = adolescent students.

Table 2

Standardised Solution for Factor Loadings for a 24-item, Six-factor Model of the POMS-A Tested on Athletes and Adolescents

Scale	Item	Sample 1 (n = 621)	Sample 2 (n = 656)	Sample 3 (n = 676)	Sample 4 (n = 596)
Anger					
	Angry	.735	.881	.814	.765
	Annoyed	.810	.789	.829	.753
	Bad-tempered	.609	.763	.671	.722
	Bitter	.647	.782	.634	.619
Confusion					
	Confused	.631	.725	.641	.605
	Mixed-up	.737	.839	.775	.794
	Muddled	.768	.837	.766	.815
	Uncertain	.639	.718	.616	.702
Depression					
	Depressed	.800	.840	.841	.761
	Downhearted	.791	.838	.719	.795
	Miserable	.591	.715	.775	.755
	Unhappy	.762	.854	.766	.764
Fatigue					
	Exhausted	.838	.886	.893	.864
	Sleepy	.715	.743	.544	.530
	Tired	.471	.811	.645	.649
	Worn-out	.848	.898	.863	.912
Tension					
	Anxious	.792	.756	.456	.594
	Nervous	.768	.669	.736	.828
	Panicky	.652	.506	.713	.706

	Worried	.814	.841	.798	.833
Vigour					
	Active	.865	.846	.777	.815
	Alert	.753	.594	.509	.489
	Energetic	.875	.899	.840	.797
	Lively	.725	.694	.652	.688

Note. All factor loadings are significant at $p < .01$.

Sample 1 = adult athletes, Sample 2 = adult students, Sample 3 = adolescent athletes, Sample 4 = adolescent students.

Table 3

Correlations Coefficients Among POMS-A Scales in Adult Athletes (n = 621), Adult Students (n = 656), Adolescent Athletes (n = 676), and Adolescent Students (n = 596)

Scale	Anger	Confusion	Depression	Fatigue
Confusion				
Adult Athletes	.490			
Adult Students	.578			
Adolescent Athletes	.613			
Adolescent Students	.587			
Depression				
Adult Athletes	.653	.631		
Adult Students	.736	.725		
Adolescent Athletes	.804	.730		
Adolescent Students	.786	.700		
Fatigue				
Adult Athletes	.387	.363	.503	
Adult Students	.295	.406	.446	
Adolescent Athletes	.273	.299	.372	
Adolescent Students	.356	.387	.275	
Tension				
Adult Athletes	.228	.561	.317	.220
Adult Students	.571	.855	.707	.413
Adolescent Athletes	.316	.512	.444	.196
Adolescent Students	.174	.492	.292	.182
Vigour				
Adult Athletes	x	x	-.220	-.351
Adult Students	x	x	-.115	-.273
Adolescent Athletes	x	x	-.149	-.336
Adolescent Students	x	x	-.079	-.276

Table 4

Multi-sample Confirmatory Factor Analysis of the POMS-A

Fit Statistics	Model 1 (df = 960)	Model 2 (df = 1014)	Model 3 (df = 1050)
χ^2	3156.799	3623.700	3966.494
χ^2:df ratio	3.288	3.570	3.778
CFI	.928	.915	.905
TLI	.917	.907	.900
RMSEA	.030	.032	.033

Note. All χ^2 values are significant at $p < .01$.

Model 1 = unconstrained; Model 2 = equal factor loadings; Model 3 = equal factor loadings and inter-correlations among mood dimensions.

Table 5

Correlations between POMS-A Scales and Criterion Measures

	POMS-A					
Criterion Measure	Anger	Confusion	Depression	Fatigue	Tension	Vigour
POMS Anger	**.89***	.50*	.64*	.49*	.38*	-.10
POMS Confusion	.48*	**.78***	.63*	.53*	.64*	-.17
POMS Depression	.64*	.65*	**.88***	.51*	.75*	-.28*
POMS Fatigue	.36*	.51*	.41*	**.90***	.40*	-.15
POMS Tension	.32*	.55*	.62*	.43*	**.76***	-.24*
POMS Vigour	-.15	-.24*	-.26*	-.26*	-.12	**.67***
STAXI Anger	**.73***	.35*	.32*	.09	.32*	.34*
PANAS Positive	-.15	.18	-.21	-.22	.04	.78*
PANAS Negative	.67*	.80*	.66*	.58*	.80*	-.03
HADS Depression	.26*	.29*	**.57***	.17	.23	-.18

* $p < .01$